

Open Research Online

The Open University's repository of research publications and other research outputs

A dynamic network analysis of emergent grammar

Journal Item

How to cite:

Ibbotson, Paul; Salnikov, Vsevolod and Walker, Richard (2019). A dynamic network analysis of emergent grammar. *First Language*, 39(6) pp. 652–680.

For guidance on citations see [FAQs](#).

© 2019 Paul Ibbotson; 2019 Vsevolod Salnikov; 2019 Richard Walker



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1177/0142723719869562>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

A dynamic network analysis of emergent grammar

Paul Ibbotson

The Open University, UK

Vsevolod Salnikov

University of Namur, Belgium

Richard Walker

The Open University, UK

Abstract

For languages to survive as complex cultural systems, they need to be learnable. According to traditional approaches, learning is made possible by constraining the degrees of freedom in advance of experience and by the construction of complex structure during development. This article explores a third contributor to complexity: namely, the extent to which syntactic structure can be an emergent property of how simpler entities – words – interact with one another. The authors found that when naturalistic child directed speech was instantiated in a dynamic network, communities formed around words that were more densely connected with other words than they were with the rest of the network. This process is designed to mirror what we know about distributional patterns in natural language: namely, the network communities represented the syntactic hubs of semi-formulaic slot-and-frame patterns, characteristic of early speech. The network itself was blind to grammatical information and its organization reflected (a) the frequency of using a word and (b) the probabilities of transitioning from one word to another. The authors show that grammatical patterns in the input disassociate by community structure in the emergent network. These communities provide coherent hubs which could be a reliable source of syntactic information for the learner. These initial findings are presented here as proof-of-concept in the hope that other researchers will explore the possibilities and limitations of this

Corresponding author:

Paul Ibbotson, Childhood, Youth and Sports Group, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

Email: paul.ibbotson@open.ac.uk

approach on a larger scale and with more languages. The implications of a dynamic network approach are discussed for the learnability burden and the development of an adult-like grammar.

Keywords

Emergence, grammar, network structure

Language is a complex adaptive system (CAS). It has many interdependent parts whose interactions and dependencies generate emergent behaviour – such as the self-organization of language over generations and the feedback loops between caregiver and child – that is difficult to model from knowledge of the parts themselves. Unlike many other complex systems in the natural world it is also a cultural one and that means, every generation, a language must be compressed through the cognitive bottleneck of what is learnable. The average toddler is only just starting to string two words together, so at this point in the cultural transmission process the structural complexity of language is reduced to almost zero. In the face of this data compression, it begs the question of how languages have evolved to be complex and how they remain so. The two major theoretical responses addressing this problem have said that either some of the complexity is organized advance of experience (Chomsky, 1957, 1965) or that the complexity is actively constructed by the child (Bybee, 2010; Croft, 2001; Givon, 1995; Goldberg, 2005; Langacker, 1987, 1991; Tomasello, 2003). There is a third way to approach this question, compatible with the orthodox dichotomy, that states that some of the complexity in language is an emergent property of many simpler entities interacting with one another, each one of which *is* learnable. In this way, a language can become more complex than is learnable (e.g. Hopper, 1998).

Because language is composed of many agents whose interactions are driven by underlying nonlinear processes, the behaviour that emerges from this is best described in probabilistic terms (Beckner et al., 2009; Holland, 1995, 1998). Learning mechanisms that incorporate this probabilistic nature into their models have successfully simulated word segmentation and phoneme discrimination (Kuhl, 2000, 2004; Saffran, Newport, & Aslin, 1996). However, the significance of this process for structural aspects of language, syntax in particular, is much more controversial (Pearl & Lidz, 2010; Pinker, 1979; Wexler & Culicover, 1983). This is partly due to the relatively abstract nature of syntax in comparison with words and sounds. Children have syntax, but they do not hear it: what they hear are utterances. The question this begs is one of developmental mechanisms: how do children get to syntax from utterances? Ever since these questions became topics of serious scientific enquiry, formal tools and computational models have provided a useful complement and challenge to experimental findings, in addition to the greater descriptive rigour and theoretical insight that the models themselves can offer (Chomsky, 1975; Pinker, 1979).

For example, connectionist-based models, based on parallel systems of artificial neurons, have had success in identifying word boundaries from sequences of phonemes, word classes from sequences of words, and phrase structure and lexical semantics from large usage corpora (Borovsky & Elman, 2006; Christiansen & Chater, 2001; Elman, 1990, 1993, 2005; Elman et al., 1996; McClelland & Rumelhart, 1988). Some models have been less concerned to represent any realistic analogue to human cognition and seek to tackle the learnability problem as a mathematical abstraction; for example Klein and Manning (2005) state that their solution ‘makes no claims to modelling human language acquisition’ (p. 35). Others, in common with the approach presented here, are more interested in how the latent structure of natural language interacts with plausible cognitive processing constraints, for example, explicitly modelling how the aspects of memory account for various syntactic phenomena or act as an aid to word learning (Freudenthal, Pine, Jones, & Gobet, 2015; Ibbotson, López, & McKane, 2018). In this spirit of cognitively-grounded proposals, lexical-based analyses have examined the degree to which the utterances a child produces can be traced to reliable and frequent multi-word patterns in the input (Bannard, Lieven, & Tomasello, 2009; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Baldwin, 1997; Lieven, Salomo, & Tomasello, 2009). In their model, Alishahi and Stevenson (2008) found that constructions gradually emerged through the clustering of different verb-frames as the model learned verb classes and constructions from artificial corpora. A major contribution to this approach was provided by McCauley and Christiansen (2019). Their model essentially tested the idea that the discovery and on-line use of multi-word units – stored in a ‘chunkatory’ – forms the basis for children’s early comprehension and production. High performance was achieved across a large number of different corpora and multiple languages, including capturing many of the features of children’s production of complex sentence types. They conclude that the model supports the idea that children’s early language can be characterized by item-based learning supported by on-line processing of distributional cues.

Language processing models have varied in the extent to which they have attempted to incorporate semantic information, with some using a supervised neural network to identify the thematic roles associated with words in sentences (Kawamoto & McClelland, 1987) while others have used a broader range of cues, including animacy, sentence position and the total number of nouns in a sentence to classify nouns as agents or patients (Connor, Gertner, Fisher, & Roth, 2008, 2009). Another approach is to consider all the possible structures given in a training corpus, and estimate their likelihood from the data (Bod, Sima’an, & Scha, 2003). This estimate can then be used to assign a structure to a new utterance by combining sub-trees from the training corpus. In its unsupervised version, this method initially assigns all possible unlabelled binary trees to an un-annotated training set, and then employs a probabilistic model to determine the most likely tree for a new utterance (Bod, 2007). In summary, a range of different theoretical models suggest categories can be recovered by distributional data, whether that is via minimum-description length clustering (Cartwright & Brent, 1997), clustering based on frequent contexts (Mintz, 2003; Mintz, Newport, & Bever, 2002), or Bayesian approaches (Griffiths & Goldwater, 2007; Parisien, Fazly, & Stevenson, 2010).

What we offer here is a much simpler methodology than many of the models reviewed above (e.g. McCauley & Christiansen, 2019), yet in common with many of them we too make use of the notion that transitional probabilities are important cues for syntactic boundaries. For example, dips in transitional probability profiles represent likely phrase boundaries and peaks indicate likely groupings of words (e.g. Thompson & Newport, 2007). Where our approach differs to others, particularly those of a connectionist or neural network orientation, is that patterns are recovered from child directed speech (CDS) unsupervised and with no a priori constraint on the number of hidden layers relevant for the particular learning task. Neither does our approach call for any specific learning biases of word learning models (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1994) other than the general capacity to represent words and the transitions between them, and to cluster frequently co-occurring words together. While acknowledging that semantic information plays an important role in construction formation, this is not formalized into our network as we want to purely assess the contribution that distributional properties make to recovering grammatical categories and the dependencies between them. Our approach also uses naturalistic CDS – the raw input out of which children construct their language – as the input to the model (cf. Reeder, Newport, & Aslin, 2013). The most important aspect of our model is that it offers a representational format that minimally departs from that of both language and the brain – namely a network of interrelated, weighted connections, whose structure evolves over time.

Incremental growth of the network captures something fundamentally developmental and complex (in the sense of many interacting parts) about the process of language acquisition, that neither batch-processing of corpus data nor non-dynamic models of development can. The dynamic networks approach offers a highly plausible psychological medium in which to simulate cognitive processes because, like language, the brain itself is a complex dynamic network (Sporns, 2002). Network studies of complex systems have shown that real world networks, such as language, are not random, as was initially assumed (Barabási, 2002; Barabási & Albert, 1999; Watts & Strogatz, 1998). The internal structure and connectivity of the system can have a profound impact upon system dynamics (Newman, Barabási, & Watts, 2006). Conceptualizing language learning as a CAS means that language acquisition research has the potential to benefit from the analytic tools developed to understand CAS in general. The approach can also offer a unified account of various linguistic phenomena, including the probabilistic nature of linguistic behaviour; continuous change within agents and across speech communities; the emergence of grammatical regularities from the interaction of agents in language use; and stage-like transitions due to underlying nonlinear processes (Holland, 1995, 1998). One such analytic tool developed for CAS analysis is community detection in networks, where network communities form around nodes (words in our case) that are more densely connected with each other than they are with the rest of the network (more detail in the Methods). We explore this idea by instantiating a corpus of early CDS into a dynamic network. We allow the network to grow word by word as the mother talks to her child, as recorded in a corpus of naturalistic speech. By using CDS, we are interested to know whether organizational properties of the

network (i.e. community structure) map onto grammatical patterns in any way that a child could plausibly capitalize on when constructing their language. If such a mapping exists, then community detection could be an important learning mechanism for the child, assuming learners sample words from the input they receive – something they presumably must do as they do not know which language community they are going to be born into.

The network we use is blind to grammatical information and its organization emerges from (a) the frequency of using a word and (b) the probabilities of transitioning from one word to another. We then implement a procedure that measures the density of links inside network communities compared to links between communities, analyse the grammatical composition of these communities and track how they develop over time. We take this approach because many decades of psycholinguistic research have shown how sensitive adults and children are to distributional patterns in language (Bloomfield, 1938; Cartwright & Brent, 1997; Finch & Chater, 1992, 1994; Goldberg, 2005; Harris, 1954; Mintz, 2003; Mintz et al., 2002; Redington, Chater, & Finch, 1998; Schütze, 1993; Tomasello, 2003).

If network communities show distinct grammatical characteristics, then the dynamic network approach suggests some of language’s complexity (grammar) can be an emergent property of how simpler elements (words) interact with one another. It would also suggest that early grammatical patterns can be represented at a level that is grounded in the distributed properties of the network.

Method

All available naturalistic CDS for two children (‘Eleanor’ and ‘Fraser’) was extracted from the Manchester Corpus (Theakston, Lieven, Pine, & Rowland, 2001). Utterances were parsed into two-word chunks (bigrams) such that ‘*John liked Mary*’ became ‘*John → liked*’ ‘*liked → Mary*’, which when implemented in a network (Figure 1) represented a total of 6861 unique words for Fraser’s CDS (displayed as nodes) and 52,057 links (or edges→) between words. For Eleanor’s CDS there were 6184 words with 65,720 links.

When Eleanor’s or Fraser’s mother said a word for the first time in the corpus a new node was added to the network. As the mothers connect two words for the first time in the corpus a new edge was added between these two nodes. As they connect the same two words as before, the weight of the edge between two nodes was increased proportionate to the frequency that this connection was made. In this manner, the network builds up distributional patterns of use. This procedure is designed to reflect what we know about distributional patterns in naturalistic corpora from other, non-network, analyses. For example, in one construction-based analysis of CDS (Cameron-Faulkner, Lieven, & Tomasello, 2003), a *What’s __X?* frame (Figure 2(a)) accounted for more than 69% of all of the CDS *What’s -is* constructions. The idea is that by instantiating these types of patterns as nodes and edges in a network, it gives community detection a way of mechanistically recovering the kinds of patterns consistent with this slot-and-frame analysis of early speech (Braine, 1987; Clark, 1974; MacWhinney, 1979). The intuition

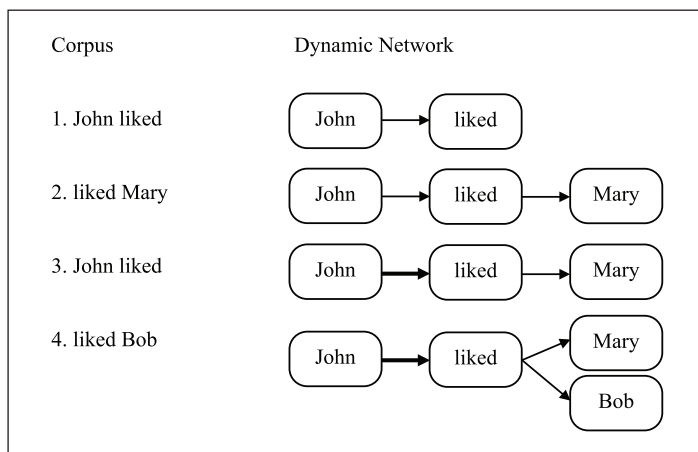


Figure 1. Schematic representation of how the network grows over time (1–4) out of words (nodes) and the relationship between those words (edges) as constructed from naturalistic speech, in this example ‘John liked Mary. John liked Bob’. Note the increased weight between repeated connections (e.g. John liked (3)).

behind the community detection algorithm is visually displayed in 2b and positioned below purposely to provide a direct comparison with the usage-generated CDS analysis of 2a (Cameron-Faulkner et al., 2003).

Informally, network communities form around nodes (words in our case) that are more densely connected with each other than they are with the rest of the network. For example, they may form around the type of *What’s_X?* collocation in Figure 2(a) or an adjective-noun phrase or noun-verb-noun pattern or any frequently co-occurring pattern or schema that is more interconnected on average than the rest of the network. Because each word (e.g. ‘dog’) has grammatical category meta-data attached to it (e.g. noun) in the corpus we could analyse the pattern of grammar not only across the whole network, but also within the communities that formed as the network developed. Importantly the network itself was blind to the grammatical information and was only built from the collocations between words that were in CDS, not their grammatical categories. A more formal description of how the model identifies categories is given in Appendix 1.

Within communities we restricted ourselves to analysing grammatical patterns of use across three-word trigrams for the practical reason that strings much longer than this became very difficult to analyse (an example of eight-word collocation flow across the whole network for Eleanor and Fraser is given in Appendix 2). Figure 3 gives a close-up of a community identified in the network of Eleanor’s CDS. From these trigram maps we characterized some of the most typical grammatical patterns for trigrams within communities, for example, the *preposition*→*determiner(article)*→*noun* for the pathway highlighted in red below.

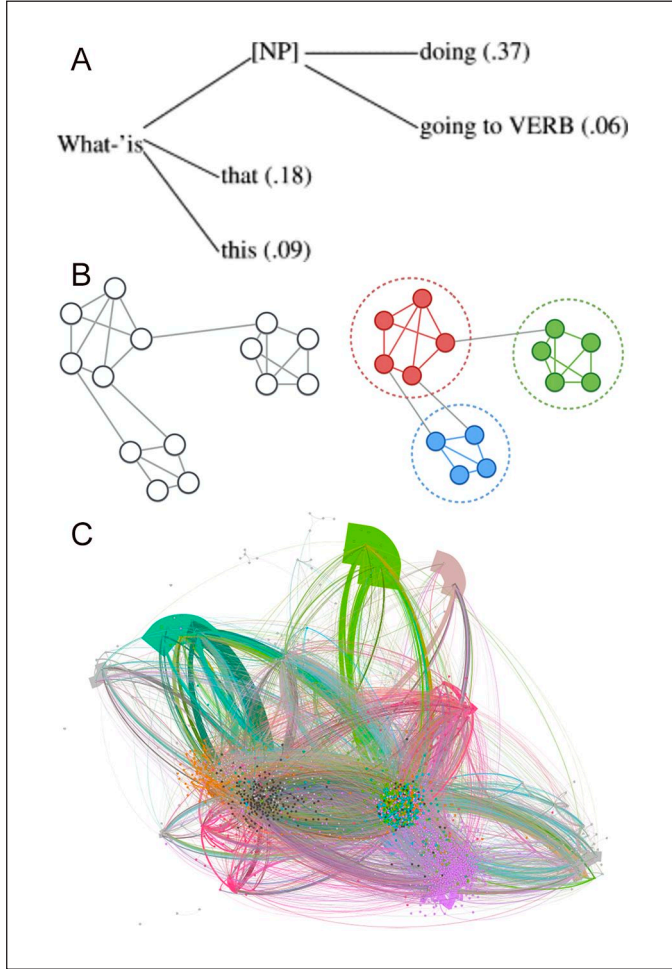


Figure 2. A. What's__X frame from Cameron-Faulkner, Lieven, and Tomasello (2003). B. Visual illustration of how communities are identified (marked by red, green and blue) around densely interconnected nodes. C. A whole network visualization of real CDS from corpus data with communities coloured. (To see this figure in colour, please view the online version of the article.)

To assess the contribution of the community detection algorithm we also ran a control procedure that essentially randomized the connections between nodes. Specifically, for each node in the final network its *in-degree* (total number of words transitioning to it) and its *out-degree* (total number of words transitioning from it) are fixed. At that point, imagine all edges in the network are cut in the middle and each node has its half-edges connected to it with the other end in the air. The process is repeated, connecting the half-edges in a random way throughout the network until there are no half-edges left to

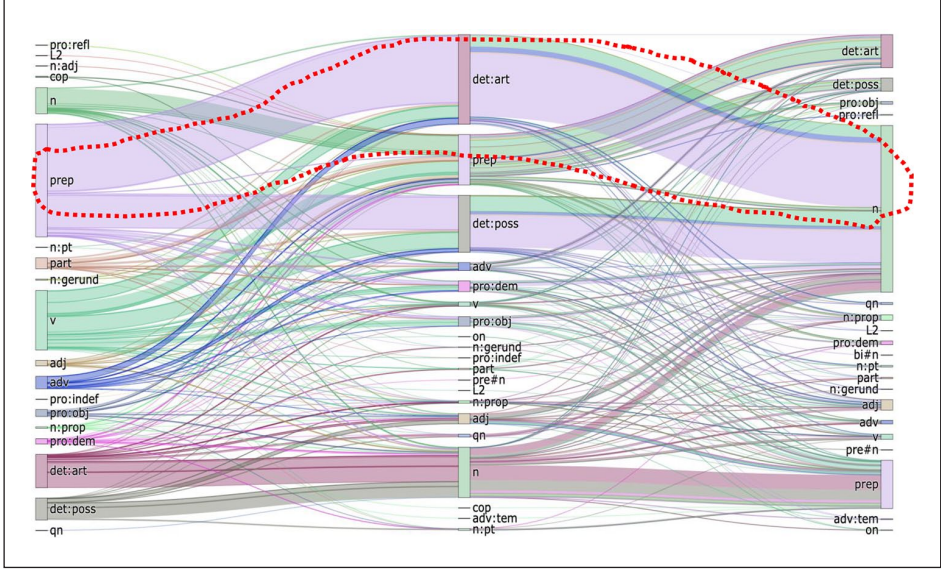


Figure 3. An example of within community trigram grammatical patterns. (To see this figure in colour, please view the online version of the article.)

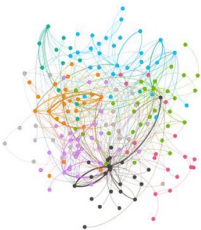
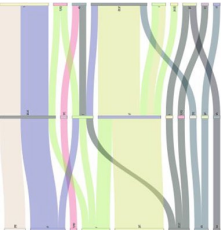
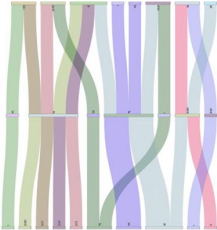
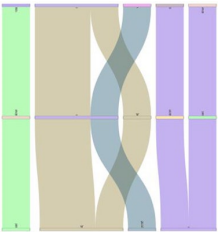
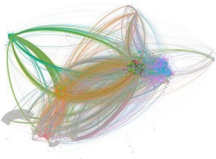
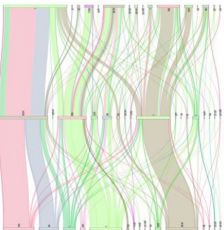
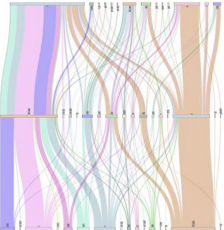
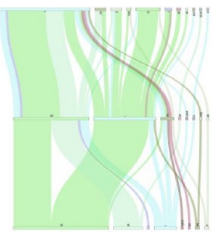
connect: take one half-edge and look randomly among all not-connected half-edges for one to connect to. In this manner the new network is a re-wired version of the old network with same nodes each having an original degree (viz. words and their neighbour connectivity), but connecting to different nodes. For dynamic networks, the timestamp of the edges (when the transition was established in the corpus of speech) can be preserved from the original distribution as well.

In what follows we present the results for the three biggest communities detected across four points of cumulative language in CDS (Tables 1 to 4). The first time point is after 200 words of the corpus, the last time point is at the last word of the corpus and two others equally divide the remaining corpus.

Our primary interest was whether frequent grammatical trigram patterns disassociate by community structure. If they do, then community structure could add valuable information which the learner could use to detect grammatical patterns.

Results

Table 1. The results of the plotting trigram grammatical flow for the three biggest communities in Eleanor’s CDS across four time points. (To see this table in colour, please view the online version of the article.).

Eleanor CDS				
#Words (# Communities)	Network visualization	Biggest community	2nd biggest community	3rd biggest community
200 (12)				
2061 (17)				

(Continued)

Table 1. (Continued)

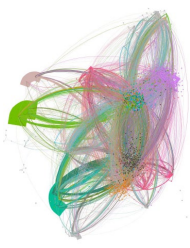
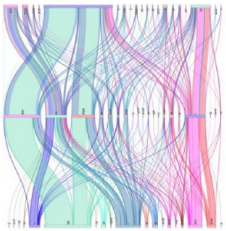
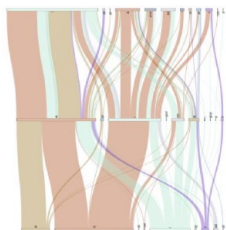
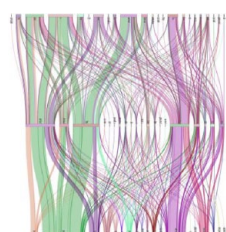
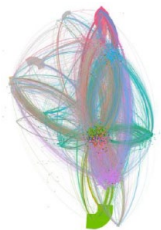
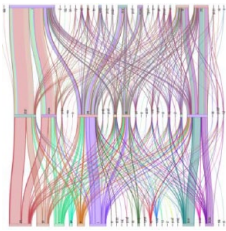
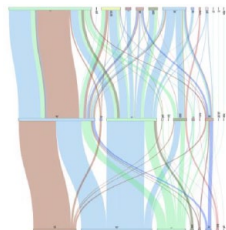
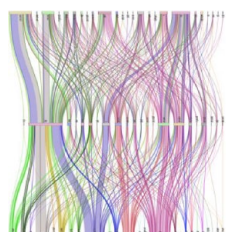
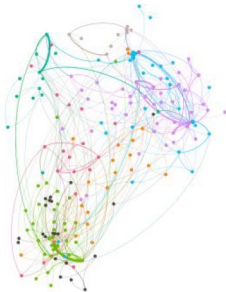
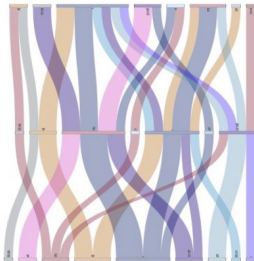
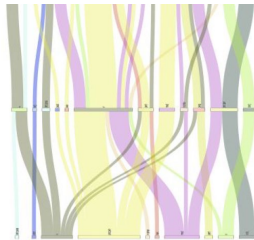
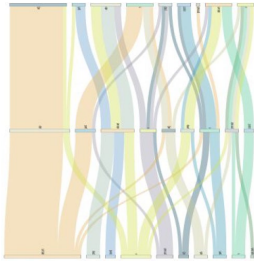
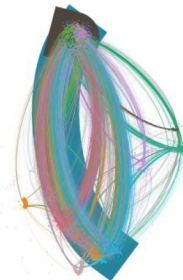
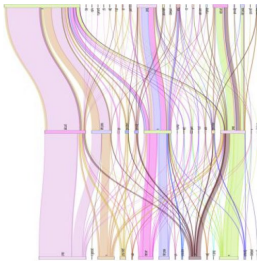
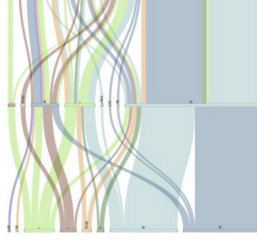
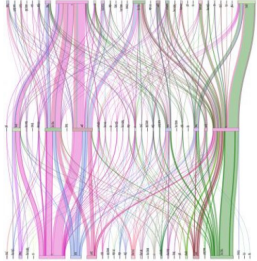
Eleanor CDS					
#Words (# Communities)	Network visualization	Biggest community	2nd biggest community	3rd biggest community	
4122 (24)					
6184 (36)					

Table 3. The results of the plotting trigram grammatical flow for the three biggest communities in Fraser's CDS across four time points. (To see this table in colour, please view the online version of the article.).

Fraser CDS				
#Words (# Communities)	Network visualization	Biggest community	2nd biggest community	3rd biggest community
200 (10)				
2287 (22)				

(Continued)

Table 3. (Continued)

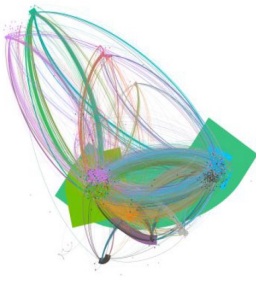
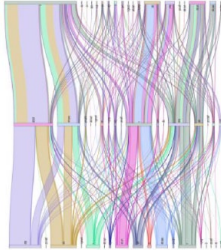
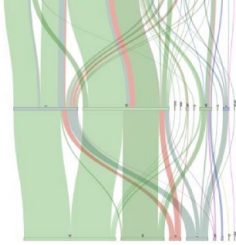
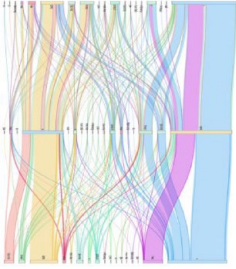
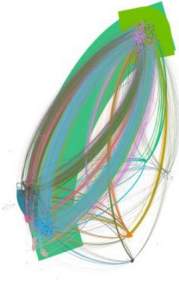
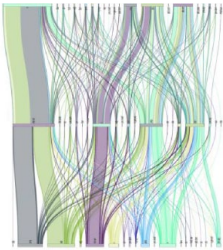
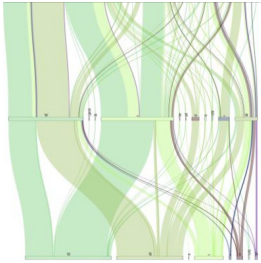
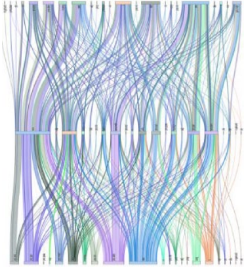
Fraser CDS				
#Words (# Communities)	Network visualization	Biggest community	2nd biggest community	3rd biggest community
4574 (22)				
6861 (24)				

Table 4. Characteristic trigram grammatical flow and lexical examples for the three biggest communities in Fraser's CDS across four time points.

Fraser CDS		2nd biggest community		3rd biggest community	
Biggest community		Examples	Grammar path	Examples	Grammar path
T1	Noun→preposition→noun	'dolly with dress' 'dog on one'	Pronoun(personal)→verb→preposition	'you hit with'	Determiner→adverb →adjective
	Noun→preposition→pronoun (indefinite)		Modal→Pronoun(personal)→verb	'can you hit'	Verb→preposition →determiner(article)
T2	Preposition→determiner(article) →noun	'with the dog'	Verb→determiner(possessive)→noun	'walk my dog'	Verb→negation→verb
	Verb→determiner(possessive)→noun	'walk your dog'	Adverb→adjective→noun	'very brown dog'	
T3	Preposition→determiner(article) →noun	'with a dog'	Adjective→adverb→noun	'brown quickly dog'	Pronoun(personal)→verb→prep osition
	Adverb→determiner(article) →noun	'slowly the dog'	Adverb→adjective→noun	'very brown dog'	Verb→preposition→noun
T4	Adverb→determiner(possessive) →noun	'slowly my dog'	Adjective→adjective→noun	'brown quickly dog'	Pronoun(subject)→verb→prepo sition
	Preposition→determiner(article) →noun	'with a dog'	Adjective→adjective→noun	'brown furry dog'	
	Adverb→determiner(article) →noun	'slowly the dog'	Adverb→adjective→noun	'very brown dog'	Pronoun(personal)→auxiliary→verb
	Determiner(article)→noun→ adverb	'the dog quickly'	Adjective→adjective→noun	'brown furry dog'	Auxiliary→pronoun(demonstrative) →pronoun(indefinite)
			Adjective→noun→adverb(temporal)	'brown dog then'	

Discussion

To recap, our primary interest concerned whether frequent grammatical (trigram) patterns in the input disassociate by community structure. If they do, then community structure represents an emergent source of grammatical information available to the language learner that is the byproduct of instantiating words and their connections into a dynamic network. First, we give a general overview of how the networks developed in line with our expectations – as a sense-check that the community network methodology works and is able to replicate previous findings – and then go on to examine the novel contribution this article makes, namely, characterizing grammatical patterns within network communities.

As one would expect, by definition of the methodology, the networks become larger and more interconnected over time, as evidenced by the increasing number of nodes and edges as the network grows (a generic feature of dynamic networks widely noted in Banavar, Maritan, & Rinaldo, 1999; Borge-Holthoefer & Arenas, 2010; Cancho & Solé, 2001; Collins & Loftus, 1975; Collins & Quillian, 1969; Hills, Maouene, Maouene, Sheya, & Smith, 2009; Steyvers & Tenenbaum, 2005). Our word-based network displays similar properties to those that focused on semantic relationships (Borge-Holthoefer & Arenas, 2010; Cancho & Solé, 2001; Collins & Loftus, 1975; Collins & Quillian, 1969; Steyvers & Tenenbaum, 2005): namely, that there are several tightly interconnected clusters with some nodes acting as bridges or hubs to other densely connected clusters. For example, Cancho and Solé (2001) demonstrated that human language displays the so-called small-world effect where the average minimum distance between two words is approximately 2–3 links, despite the fact there are many thousands of words in the language network. This is possible because not all nodes in the network are created equal – there are some hub nodes that are much more interconnected than others. With this combination of local structure combined with global access, these networks become increasingly small-world and approximate a structure that is thought to aid efficient language processing and production (Banavar et al., 1999; Hills et al., 2009) and even account for some differences between early and late talkers (Beckage, Smith, & Hills, 2011). For example, Beckage et al. (2011) showed that the networks of typically-developing children show small-world structure as early as 15 months and with as few as 55 words in their vocabulary. By contrast, children with language delay display this structure to a lesser degree, causing a maladaptive bias in word acquisition for late talkers, potentially indicating a preference for infrequent words. The fact that there is this small-world non-uniform distribution of connectivity allows the community detection algorithm to identify clusters of densely interconnected nodes.

From the grammatical analysis in Tables 2 and 4 of characteristic pathways through communities, it appears that communities are able to differentiate patterns of grammatical use in CDS. For example, for Fraser’s CDS at time point 1, Community 1 contained a *Noun*→*preposition*→*noun* frequent pathway, Community 2 *Pronoun(personal)*→*verb*→*preposition* and Community 3 *Determiner*→*adverb*→*adjective*. For Eleanor’s CDS at time point 4, Community 1 identified *Preposition*→*determiner(article)*→*noun*, Community 2 *Adjective*→*noun*→*adverb* and Community 3 *Pronoun(interrogative)*→*Copula*→*Pronoun(demonstrative)*. Within this general picture, there are interesting individual differences in developmental patterns. For example, for Eleanor’s CDS, the

complexity of the plots varies by community size, with the largest communities and smallest communities becoming progressively more complex with each epoch, while the second largest community becomes progressively more skewed, with a small number of strong pathways and a large number of small pathways. For Fraser’s CDS, different patterns emerge. The largest community in the final epoch appears more skewed than in the penultimate epoch. The cross-tabulations in Figure 4 summarize the pattern in Tables 2 and 4, showing how the characteristic grammatical pathways disassociate by community structure; that is, one community structure has a different grammatical hub in comparison with another. The graph in Figure 4 also shows how the control procedure eventually works to undermine community structure, in contrast to the natural community structure growth inherent in language. At the start of the network building there is a period where the control procedure generates more community structure than natural language, and the reason is evident from Figure 5. At the beginning the control procedure has lots of sub-communities which are small but meaningless. As the network grows these are subsumed into an ever longer but meaningless string of connected words that is captured by fewer and fewer communities. Natural language shows the opposite pattern, with sustained growth in community structure. We know that children are sensitive to the kinds of distributional patterns the network instantiates, so it seems plausible that community structure could provide an emergent source of information for the learner when constructing their early grammar. Moreover the patterns within communities contained some of the basic grammatical building blocks of English: verbs, nouns, adverbs, adjectives, modals, auxiliaries and determiners. So, the patterns that are constructed from these units could provide a foothold into the basic *who-did-what-to-whom* that a grammar organizes. Because the grammatical network was tagged with grammatical categories and had the ability to represent transitions between those categories, communities can contain ordered lexical class templates, for example a noun-verb-noun schema able to represent ‘dog bites man’ and ‘man bites dog’. Grammatical generalizations at an abstraction higher than that level (e.g. subject- or agenthood) were not examined here, although there is no reason why the same methodology of community detection could not be applied to corpora tagged with that data – the question would be the same in that instance, namely whether community structure can be dissociated by subjecthood, for example. We chose syntactic classes as they represent the least abstracted level away from the words themselves and presumably a level over which ever more abstract categories are later generalized (see suggestion later in the Discussion for hierarchical community detection).

Because of some of the analytic complexities involved we focused on the three biggest communities. Just from looking at these three communities though, it is clear that although there is dissociation across communities at any one time point there is also a fluid characterization of their grammatical pathways within communities across time. For example, as Tables 2 and 4 show, the same grammatical pattern is not always the most characteristic of that community across all four time points. To some extent this is to be expected as patterns from smaller communities get subsumed into larger communities as the network grows. Clearly, we need to know more about all the communities for a fuller picture, including whether the communities begin to converge on individual stable grammatical identities after a period of input saturation – something that further investigation could confirm over a larger corpus.

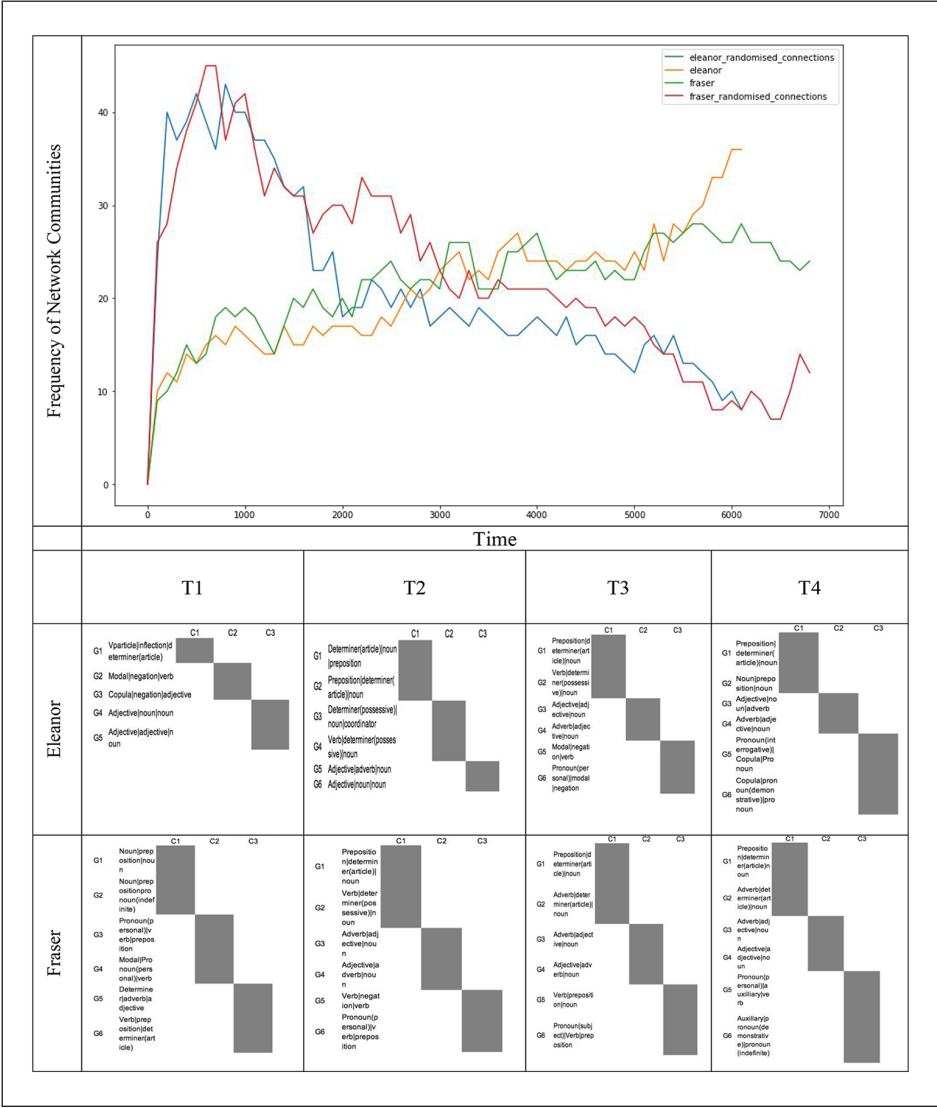


Figure 4. The graph at the top shows community development for the two networks and their associated randomized controls over time. Beneath the graph is a visual summary of the data in Tables 2 and 4 presented as cross-tabulations and which highlights the dissociation between community (C1–3) and grammatical pattern (G1–6). (To see this figure in colour, please view the online version of the article.)

What we hope to offer here is a modest proof-of-concept for the community structure approach to grammatical pattern identification and the potential insight that the dynamic network approach can offer. We encourage other researchers to explore the possibilities

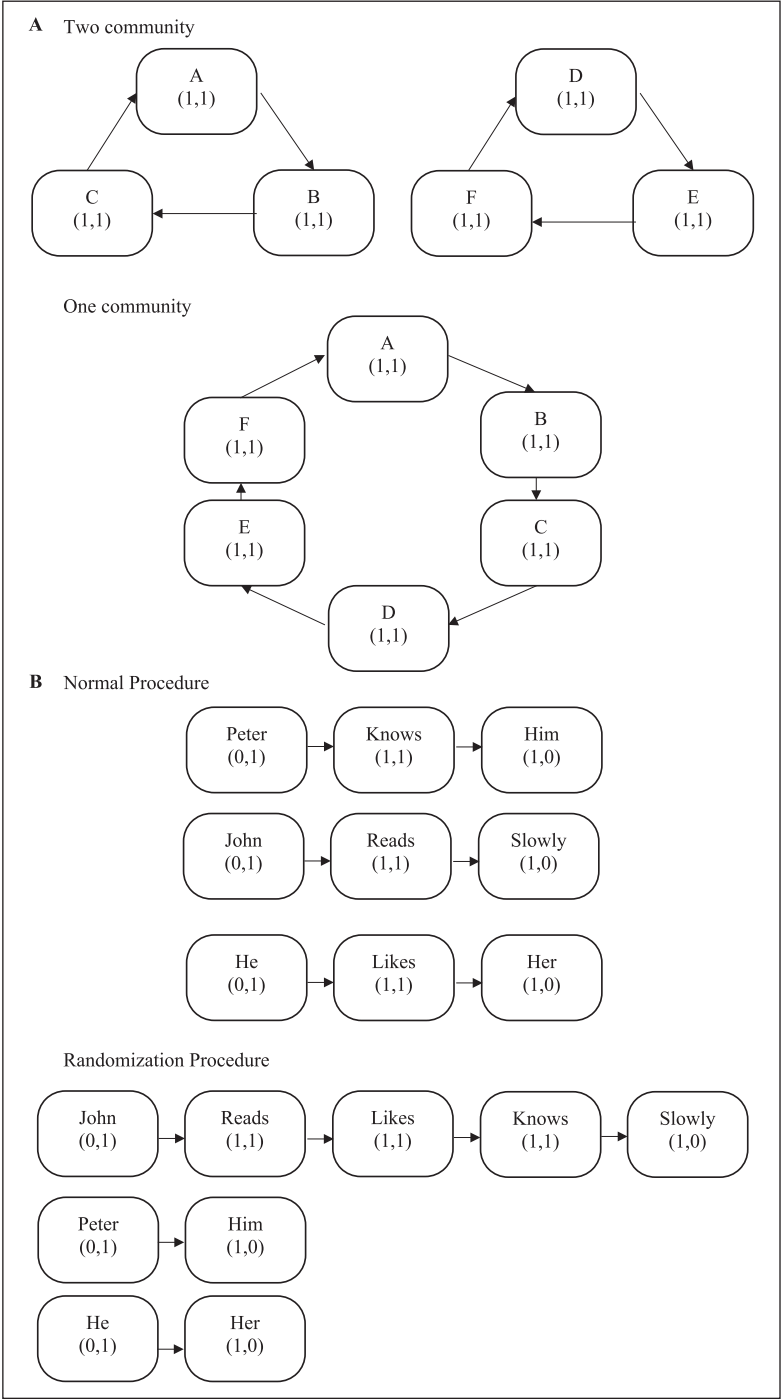


Figure 5. (Continued)

Figure 5. A. Two hypothetical networks with an identical number of nodes, each with the same in-degree ($x,0$) and out-degree ($0,y$) but that are most naturally clustered into two communities and one community. Communities are formally captured by the community detection algorithm, see Appendix I. B. The effect of the randomization procedure on the emergent network: while preserving the degree, the dynamic properties (when words are added and connections made) and the number of nodes (words and their connectivity), the randomization procedure scrambles the pathway and therefore the grammar. The result of the control procedure is a network that has the same nodes with the same in- and out-degree as the normal procedure, but whose pathways become meaningless.

and limitations of this approach on a larger scale. For example, the scope of this article was limited to analysing one language, with two input sources over four time periods and the three largest communities that emerged from that input. A more mechanistic way of recovering characteristic patterns within communities (which was done by hand here) would allow the methodology to be scaled up to cover more time points, more speakers, more communities and more languages. The cross-linguistic validity of the approach is especially important because, for example, a noun is a noun because of its position relative to its grammatical neighbours: in English, it gets modified by an adjective, follows a determiner and precedes a verb. But because the child obviously does not know it is going to be born into an English-speaking community, any learning mechanism needs to be robust enough to handle the variation in word order and other aspects of morphosyntactic variation across languages. Because the raw input into the network is words and their transitions (not grammatical categories) it is predicted that community structure detection would be able to operate with some success cross-linguistically, but this prediction obviously needs to be rigorously tested further. In comparison with previous models (e.g. McCauley & Christiansen, 2019) the architecture of the community detection approach is much simpler and so at present limits the kinds of data we are able to simulate. For example, modelling the trajectory of overgeneralization errors is currently not possible, although it is possible to see how network metrics are relevant here too. What the approach loses in its power to demonstrate productivity, it potentially gains in its cognitive plausibility: in a network of interrelated, weighted connections, whose structure evolves over time, it offers a representational format that minimally departs from that of both language and the brain. This, of course, is not a mutually exclusive offer to those more sophisticated models that integrate semantic information or a chunkatory architecture which must surely be part of a more comprehensive story of language acquisition. However, it could be argued that the contribution of this approach is to emphasize how much structure there already is in language when it is presented as a dynamic network of communities, before this information gets fed into these more sophisticated models.

One very interesting way in which the analysis here could be further developed would be to begin to layer the networks and analyse the relationships between layers. The motivation for this is that almost all linguistic theories subscribe to some level of hierarchical organization in language (although they may profoundly disagree where this hierarchy comes from and how detached it becomes from meaning). Most theories admit a role for hierarchy to escape analysis based only on form that, for example, prevents an adequate

account of long-distance dependencies or the fact the children readily interpret the novel *the gazzer mibbed the toma* as a transitive utterance despite not having had experience of the forms *gazzer*, *mibbed* and *toma*.

In theory community detection offers a way to provide such hierarchical categorization, although demonstrating this is beyond the scope of the present article. It would do this by treating the communities that emerge from the CDS (the ones established in the present study) as *the input*, or nodes to a second layer of the network in something like the schematization in Appendix 3. The idea here would be that this second layer abstracts away from the form and describes relationships between communities identified at a lower level. The extension to the methodology established here potentially has important consequences for theories that admit some role for hierarchy. For example, in usage-based approaches to language, grammar is often characterized as a structured inventory of constructions, conceptualized as some sort of organized network of linguistic form and function (Bybee, 2010; Croft, 2001; Givon, 1995; Goldberg, 2005; Langacker, 1987; Tomasello, 2003). Precisely what this inventory looks like is often not specified in any detail, and where it is, the proposals are often static, highly schematized (viz. hierarchical abstraction) and only partial visualizations of the complete grammatical system. By instantiating language in a dynamic, layered network, we would catch this inventory in the act of being built and visualize what distributional patterns of grammar use might look like for a child acquiring language. In doing so, this approach offers something more concrete, incremental and fleshes out what is meant by the theoretical construct ‘structured inventory’.

The communities detected in the CDS examined here are a byproduct of organizing language into a network that is sensitive to the frequency of word use and the transition between words – something that we know adults and children are sensitive to (Bloomfield, 1938; Braine, 1987; Cartwright & Brent, 1997; Finch & Chater, 1992, 1994; Goldberg, 2005; Harris, 1954; Mintz, 2002, 2003; Mintz et al., 2002; Redington et al., 1998; Schütze, 1993; Tomasello, 2003). It seems plausible to suggest that if this information is available to learners then they might use it as a way of beginning to categorize and organize their grammatical experience. We know language is a classic example of a complex system. It has multiple speakers interacting with one another; it is adaptive to past behaviour and a speaker’s behaviour is the consequence of many competing factors that operate on many interrelated time cycles. If some of the complexity of language can be a byproduct of complex dynamic systems, not all of it needs to be actively constructed or organized in advance of experience. That means some of the burden of learning the complexity of language can be outsourced to emergence, and the cognitive bandwidth of the children who learn it no longer places the same constraints. In this way, language can sustain intergenerational complexity far beyond the complexity of the learning mechanisms acquiring it.

A strength of the dynamic network approach is that it offers a method of representing language growth that minimally differs from the way language is actually used, and that means the gap between theoretical construct and data is kept small. It also presents a way to ground linguistic representation in a medium that is psychologically plausible; for example, the usage-based proposal that frequently occurring patterns are stored together as templates or schemas can be grounded in the

community structure of the network which in turn can be grounded in the Hebbian learning principle that neurons that fire together wire together (Hebb, 1949; Lowel & Singer, 1992). Here we formally make this link between distributional learning, the schemas of usage-based theory and the community structure of a network. We hope more researchers from across the linguistic and cognitive theoretical spectrum will find use in this method for visualizing, examining and testing theories about language development.

Author contributions

PI conceived of the idea, directed the analysis and wrote the paper. RW extracted the relevant data from the online corpora to feed into the network analysis and VS processed the network data, provided analysis of community content and created the plots. All authors critically revised the manuscript for important intellectual content.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Paul Ibbotson <https://orcid.org/0000-0001-8346-0616>

References

- Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science: A Multidisciplinary Journal*, 32, 789–834.
- Banavar, J. R., Maritan, A., & Rinaldo, A. (1999). Size and form in efficient transportation networks. *Nature*, 399, 130–132.
- Bannard, C., Lieven, E., & Tomasello, M. (2009). Early grammatical development is piecemeal and lexically specific. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 17284–17289.
- Barabási, A. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus Books.
- Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 5439, 509–512.
- Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*, 65(5), e19348.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59, 1–26.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. doi:10.1088/1742-5468/2008/10/P10008.
- Bloomfield, L. (1938). Language or ideas? *Language*, 12, 89–95.
- Bod, R. (2007). Is the end of supervised parsing in sight? In A. Zaenen, & A. van den Bosch (Eds.), *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 400–407). Stroudsburg, PA: Association for Computational Linguistics.
- Bod, R., Sima'an, K., & Scha, R. (2003). *Data-oriented parsing*. Stanford, CA: CSLI Publications.

- Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: Structure and dynamics. *Entropy*, 12, 1264–1302.
- Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33, 759–790.
- Braine, M. D. S. (1987). What is learned in acquiring word classes: A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Lawrence Erlbaum.
- Bybee, J. (2010). *Frequency of use and the organization of language: Frequency of use and the organization of language*. Oxford, UK: Oxford University Press.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843–873.
- Cancho, R. F., & Solé, R. V. (2001). The small world of human language. *Proceedings of Biological Sciences*, 268, 2261–2265.
- Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, 63, 121–170.
- Chomsky, N. (1957). *Syntactic structures*. New York, NY: Mouton de Gruyter.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1975). *The logical structure of linguistic theory*. New York, NY: Springer.
- Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, 5, 82–88.
- Clark, R. (1974). Performing without competence. *Journal of Child Language*, 1, 1–10.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Connor, M., Gertner, Y., Fisher, C., & Roth, D. (2008). Baby SRL: Modeling early language acquisition. In A. Clark & K. Toutanova (Eds.), *Proceedings of the Twelfth Conference on Computational Natural Language Learning – CONLL’ 08* (pp. 81–88). Stroudsburg, PA: Association for Computational Linguistics.
- Connor, M., Gertner, Y., Fisher, C., & Roth, D. (2009). Minimally supervised model of early language acquisition. In J. Hajič (Ed.), *Proceedings of the Thirteenth Conference on Computational Natural Language Learning – CONLL ’09* (pp. 84–92). Stroudsburg, PA: Association for Computational Linguistics.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford, UK: Oxford University Press.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences*, 9, 111–117.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: The MIT Press.
- Finch, S., & Chater, N. (1992). Bootstrapping syntactic categories using statistical methods. In D. Daelemans & W. Powers (Eds.), *Background and experiments in machine learning of natural language: Proceedings of the 1st SHOE workshop on statistical methods in natural language* (pp. 229–236). Tilburg, The Netherlands: ITK.

- Finch, S., & Chater, N. (1994). Distributional bootstrapping: From word class to proto-sentence. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 301–306). Mahwah, NJ: Lawrence Erlbaum.
- Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2015). Simulating the cross-linguistic pattern of optional infinitive errors in children’s declaratives and Wh- questions. *Cognition*, 143, 61–76.
- Givon, T. (1995). *Functionalism and grammar*. Amsterdam, The Netherlands: John Benjamins.
- Goldberg, A. (2005). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, 21, 125–155.
- Griffiths, T. L., & Goldwater, S. (2007). A fully Bayesian approach to unsupervised part-of-speech tagging. In A. Zaenen, & A. van den Bosch (Eds.), *ACL’07 – 45th Annual Meeting of the Association of Computational Linguistics* (pp. 744–751). Stroudsburg, PA: Association for Computational Linguistics.
- Harris, Z. S. (1954). Distributional structure. *WORD*, 10, 146–162.
- Hebb, D. O. (1949). *The organization of behavior*. New York, NY: John Wiley.
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, 112, 381–396.
- Holland, J. H. (1995). *Hidden order: How adaption builds complexity*. Boston, MA: Addison Wesley.
- Holland, J. H. (1998). *Emergence: From chaos to order*. Oxford, UK: Oxford University Press.
- Hopper, P. (1998). Emergent grammar. In M. Tomasello. (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 143–161). Mahwah, NJ: Lawrence Erlbaum.
- Ibbotson, P., López, D. G., & McKane, A. J. (2018). Goldilocks forgetting in cross-situational learning. *Frontiers in Psychology*, 9, 1301.
- Kawamoto, A. H., & McClelland, J. (1987). Mechanisms of sentence processing: Assigning roles to constituents of sentences. In *Parallel distributed processing. Vol. 2: Psychological and biological models* (pp. 195–248). Mahwah, NJ: Lawrence Erlbaum.
- Klein, D., & Manning, C. D. (2005). Natural language grammar induction with a generative constituent-context model. *Pattern Recognition*, 38, 1407–1419.
- Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11850–11857.
- Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.
- Langacker, R. W. (1987). *Foundations of cognitive grammar. Vol. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, R. W. (1991). *Foundations of cognitive grammar. Vol. II: Descriptive application*. Stanford, CA: Stanford University Press.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30, 333–370.
- Lieven, E., Pine, J., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219.
- Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children’s production of multi-word utterances: A usage-based analysis. *Cognitive Linguistics*, 20, 481–507.
- Lowel, S., & Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255, 209–212.

- MacWhinney, B. (1979). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, 43, 1–123.
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, 92, 199–227.
- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, 126, 1–51.
- McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA: The MIT Press.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition*, 30, 678–686.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91–117.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26, 393–424.
- Newman, M., Barabási, A., & Watts, D. (2006). *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.
- Parisien, C., Fazly, A., & Stevenson, S. (2010). An incremental bayesian model for learning syntactic categories. In A. Clark & K. Toutanova (eds.), *CoNLL '08 Proceedings of the Twelfth Conference on Computational Natural Language Learning* (pp. 89–96). Stroudsburg, PA: Association for Computational Linguistics.
- Pearl, L., & Lidz, J. (2010). When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development*, 5, 235–265.
- Pinker, S. (1979). Formal models of language learning. *Cognition*, 7, 217–283.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425–469.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66, 30–54.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- Schütze, H. (1993). Part-of-speech induction from scratch. In L. Schubert (Ed.), *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics* (pp. 251–258). Stroudsburg, PA: Association for Computational Linguistics.
- Sporns, O. (2002). Network analysis, complexity, and brain function. *Complexity*, 8, 56–60.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks. *Cognitive Science*, 29, 41–78.
- Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, 28, 127–152.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3, 1–42.
- Tomasello, M. (2003). *Constructing a language. A usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.
- Wexler, K., & Culicover, P. (1983). *Formal principles of language acquisition*. Cambridge, MA: The MIT Press.

Appendix I

Formal detail on the community detection algorithm

To determine the extent which certain nodes cluster together in the network a modularity detection algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) was run on the network for each developmental time window. Defined as a value between -1 and 1 , modularity measures the density of links inside communities compared to links between communities. For a weighted network, modularity is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

where A_{ij} represents the weight of the edge between i and j , $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex i , c_i is the community to which vertex i is assigned, the δ function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2} \sum_{ij} A_{ij}$. In order to maximize this value efficiently, the Louvain Method was applied (Blondel et al., 2008). First, each node in the network is assigned to its own community. Then for each node i , the change in modularity is calculated for removing i from its own community and moving it into the community of each neighbour j of i . This value is easily calculated by:

$$\Delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right],$$

where \sum_{in} is the sum of all the weights of the links inside the community i is moving into, \sum_{tot} is the sum of all the weights of the links to nodes in the community, k_i is the weighted degree of i , $k_{i,in}$ is the sum of the weights of the links between i and other nodes in the community, and m is the sum of the weights of all links in the network. Then, once this value is calculated for all communities i is connected to, i is placed into the community that resulted in the greatest modularity increase. If no increase is possible, i remains in its original community. This process is applied repeatedly and sequentially to all nodes until no modularity increase can occur. Once this local maximum of modularity is hit, the first phase has ended.

The second phase of the algorithm groups all of the nodes in the same community and builds a new network where nodes are the communities from the previous phase. Any links between nodes of the same community are now represented by self-loops on the new community node, and links from multiple nodes in the same community to a node in a different community are represented by weighted edges between communities. Once the new network is created, the second phase has ended and the first phase is reapplied to the new network.

Appendix 2

Eight-word templates

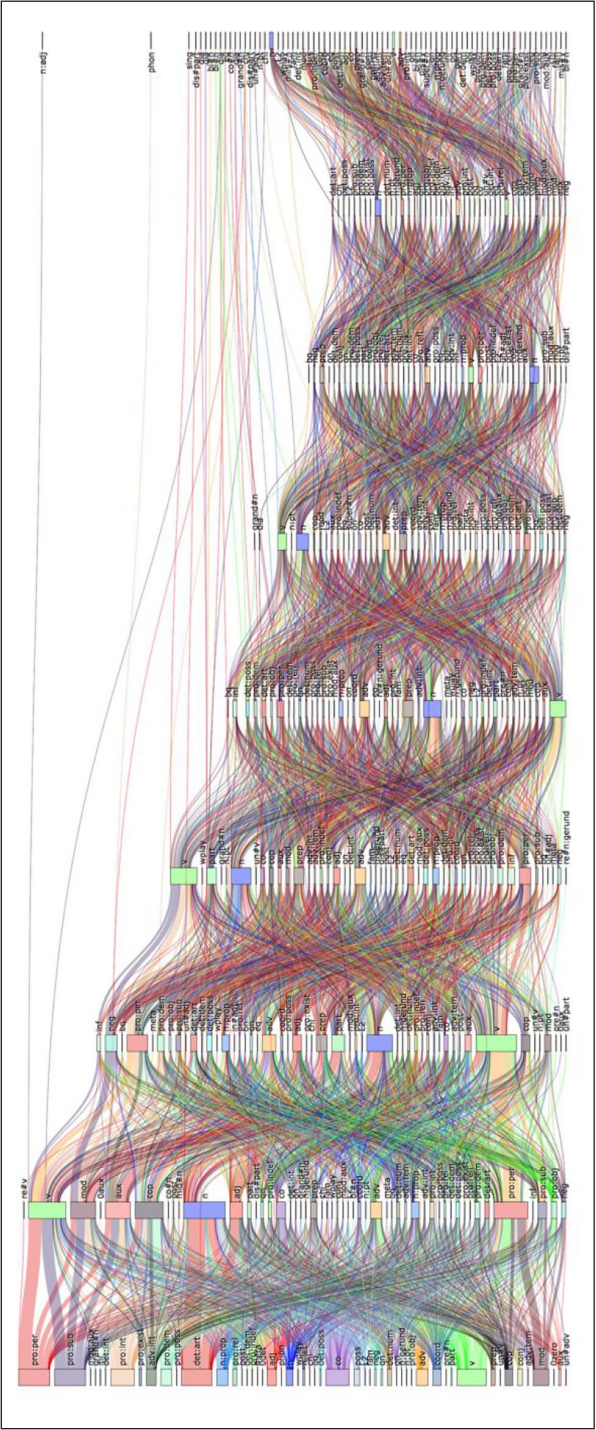


Figure 6. An eight-word template for grammatical flow for Eleanor’s entire CDS network. (To see this figure in colour, please view the online version of the article.)

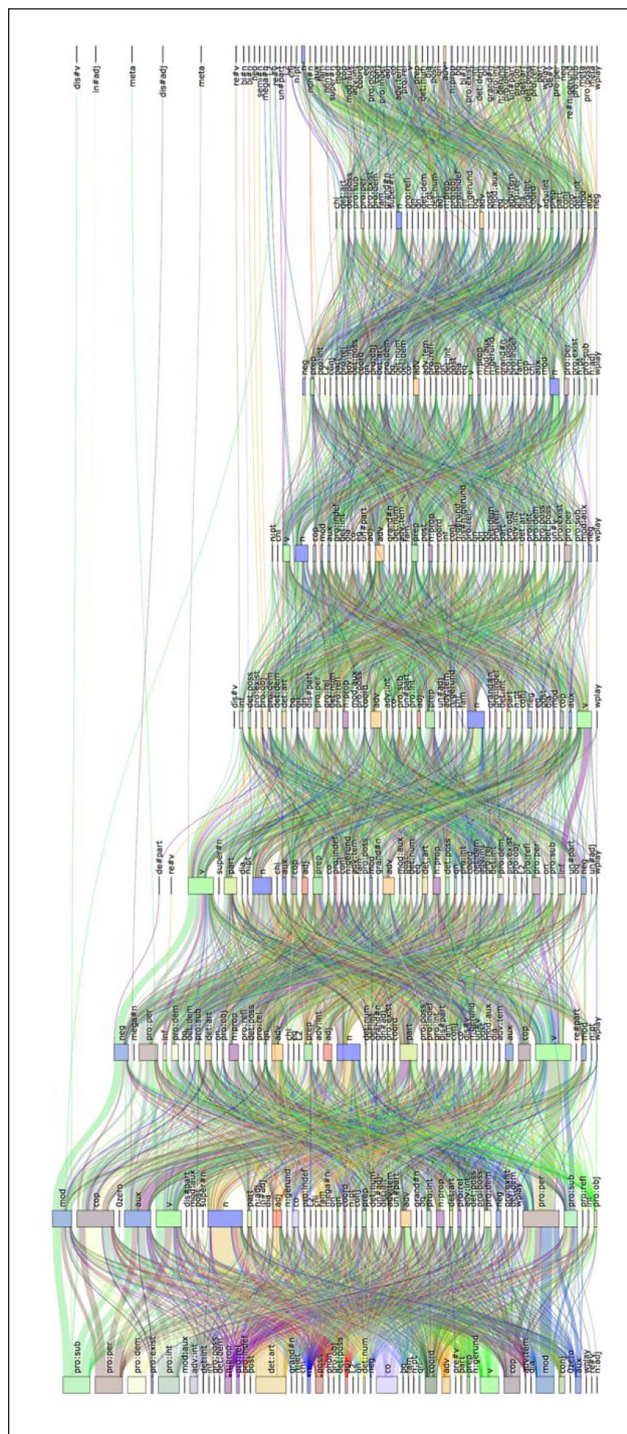
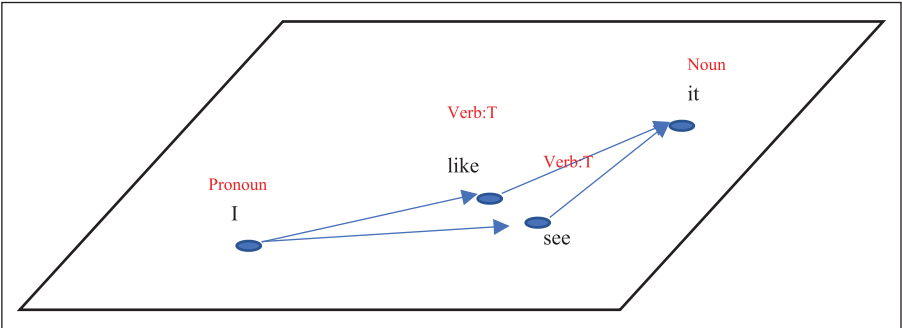


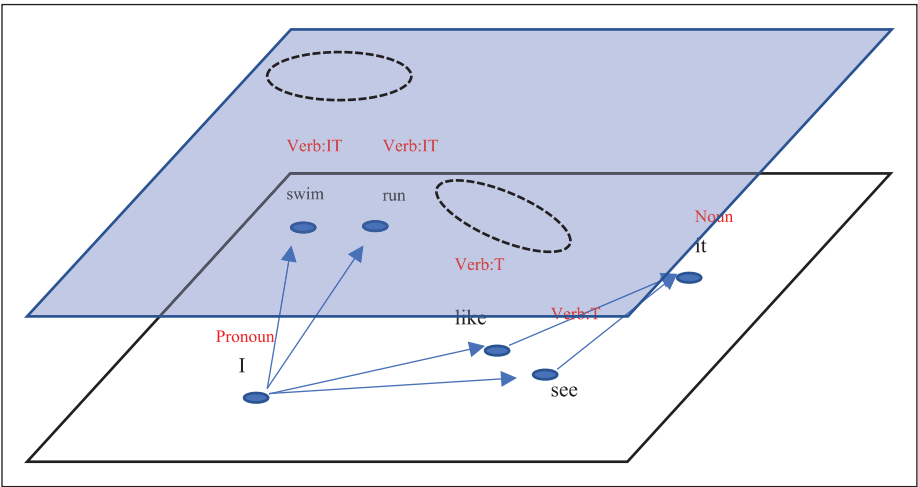
Figure 7. An eight-word template for grammatical flow for Fraser's entire CDS network. (To see this figure in colour, please view the online version of the article.)

Appendix 3

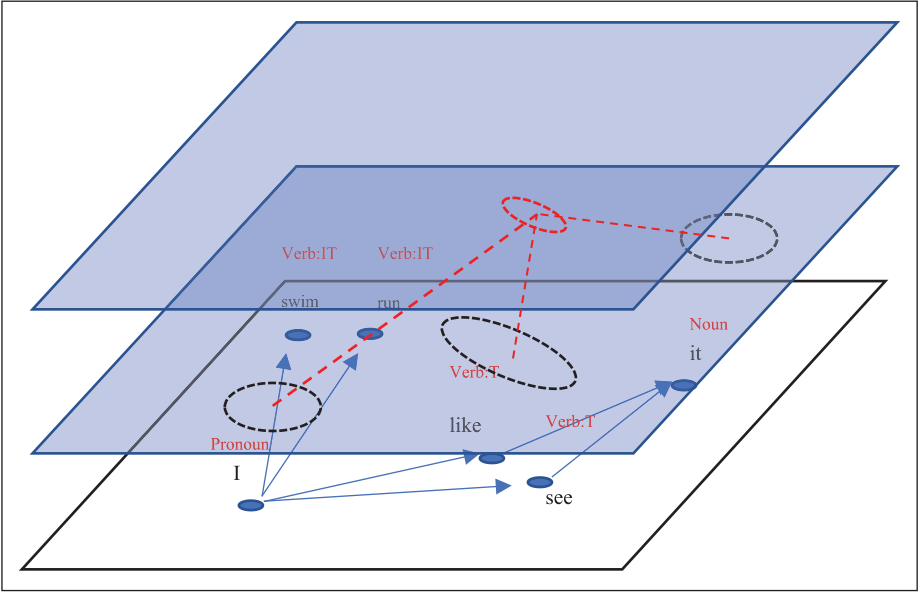
A layered community detection procedure



Time 1. Words populate the network. Grammatical tags are associated with the words but not available to the network.



Time 2. Communities form around grammatical patterns.



Time 3. Community detection at a lower level treated as input nodes for the higher level. Community detection at this higher level might be able to detect meaningful paths between lower level communities to form an organized hierarchy.

(To see these diagrams in colour, please view the online version of the article.)